SOTI

10 YEARS
OF SECURITY INSIGHT

V10 EDIÇÃO 03

Extraindo seus resultados:

como os web scrapers impactam o comércio eletrônico

Akamai

State of the Internet/Security

Índice

Você sabia que os bots geram mais da metade de todo o tráfego da Web? O setor de comércio, em particular, com sua dependência de aplicativos e ativos da Web que geram receita, tem sido mais afetado pelo tráfego de bots de alto risco (Figura 1). E, embora muitas vezes ouçamos que os bots estão evoluindo, os bots web scrapers são o tipo que chama a atenção das organizações orientadas para o comércio eletrônico hoje em dia, pois seus impactos econômicos, muitas vezes ocultos abaixo da superfície, diferem de outros tipos de bots. A detecção de bots scrapers também se tornou muito mais difícil devido ao aumento de botnets de IA (inteligência artificial) e tecnologias de navegadores headless, o que os torna extremamente evasivos. Por exemplo, um dos clientes da Akamai do setor de comércio eletrônico teve 99% do tráfego de alto risco interrompido, e nem sabia que esse tráfego era de bots scrapers.

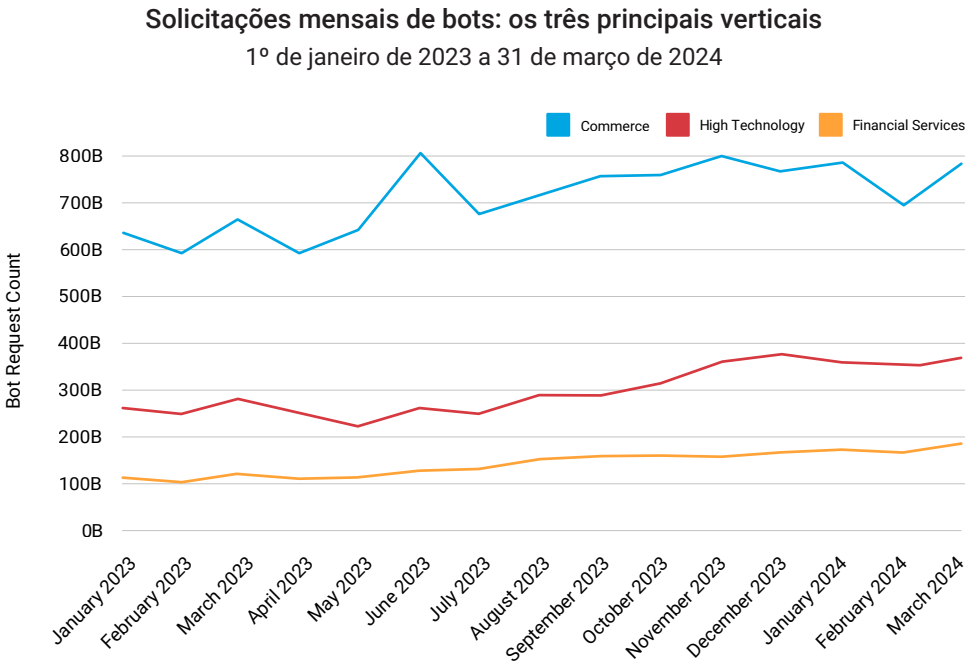


Fig. 1: O comércio é o principal vertical para solicitações de bots, e um aumento no tráfego global de bots no setor de comércio pode ser observado do início de 2023 até o primeiro trimestre de 2024

Portanto, neste relatório State of the Internet (SOTI), focamos na evolução e especialização desses bots (e de seus operadores). Embora os bots estejam por aí há algum tempo, continuamos a ver sua aplicação em vários grupos para permitir ataques criminosos, esquemas de fraude e inteligência competitiva. Recentemente, vimos uma tendência de aumento do uso de todos os bots e um aumento nos impactos negativos dos bots scrapers nos negócios. Este relatório foi elaborado para compartilhar insights técnicos e metodologia de ataque para aumentar a conscientização sobre esse problema crescente em todo o setor de comércio.

Bots: o bom, o mau e o feio

Toda organização focada em comércio eletrônico sofre com bots que estão em constante evolução e se tornando mais especializados, dependendo do que pretendem realizar. No setor de comércio, há uma grande variedade de tipos de bots que executam muitas tarefas diferentes. Uma maneira fácil de pensar neles é dividi-los em três grupos: bots bons, bots mal-intencionados e bots cinza. Os bots bons ajudam os clientes a encontrar seu website. Bots mal-intencionados capturam seu website para fins mal-intencionados. Os bots cinza tendem a ser barulhentos, embora ainda sejam legítimos; eles são uma subcategoria dos bots bons (por exemplo, bots de parceiros que estão constantemente fazendo ping e outras APIs de programas que fazem chamadas frequentes).

Assim, à medida que pensamos em chatbots úteis e bots de mecanismos de pesquisa que podem ter impactos benéficos, como responder às perguntas básicas dos usuários e fornecer conteúdo de website que retorne resultados de pesquisa mais precisos, queremos otimizar esses tipos de bots, ao mesmo tempo em que contemos os custos de TI. Para os prejudiciais, como os bots de preenchimento de credencial que tentam obter acesso não autorizado à conta de um cliente, levando à apropriação indevida de conta, queremos tomar medidas preventivas sem afetar a experiência geral do cliente. Um tipo de bot que atingiu o cenário recentemente está se tornando especialmente problemático ao reduzir a receita, diminuir a fidelidade e aumentar os custos: os bots web scrapers.

Os bots scrapers, um botnet usado para extrair diretamente dados e conteúdo de websites na Internet, são únicos. Eles estão chamando a atenção por causa de como operam de forma diferente e de como seus impactos nos negócios e detecções são diferentes em relação a outros bots. Web scrapers também são multifacetados, pois seus casos de uso variam dependendo de como as organizações e os operadores monetizam as informações coletadas por esses bots. Independentemente do objetivo específico, os scrapers estão custando receita, aumentando os custos de TI e afetando negativamente as experiências gerais dos clientes.

Neste relatório SOTI, examinamos os impactos da raspagem em todo o comércio eletrônico e examinamos por que os proprietários de empresas (como dos setores digital, marketing, marca, finanças, risco e segurança) devem ter um interesse compartilhado em impedir scrapers abusivos. Para entender melhor esses impactos, é fundamental visualizar o panorama completo do motivo pelo qual os bots web scrapers evoluíram, para o que estão sendo usados, como operam, quais são seus impactos e o que as organizações de comércio podem fazer em relação a eles.

Principais insights do relatório

- O web scraping não é apenas um problema de fraude ou segurança, mas também um problema de negócios. Os bots scrapers têm um efeito negativo em muitas facetas da organização, incluindo receita, vantagem competitiva, identidade da marca, experiência do cliente, custos de infraestrutura e experiência digital, apenas para citar alguns.

- De acordo com um estudo de caso de pesquisa da Akamai, 42,1% da atividade geral de tráfego era proveniente de bots, com 65,3% desse tráfego sendo de bots mal-intencionados. E um total de 63,1% do tráfego de bots mal-intencionados usava técnicas avançadas.

- A tecnologia de navegador headless mudou o cenário de scraper, exigindo uma abordagem para gerenciar esse tipo de atividade de bot que é mais sofisticada do que outras mitigações baseadas em JavaScript.

- Os impactos técnicos que as organizações enfrentam como resultado da raspagem, seja a raspagem feita com intenções maliciosas ou benéficas, incluem degradação do desempenho do website, poluição métrica do website, ataques a credenciais comprometidas de websites de phishing, aumento dos custos de computação e muito mais.

- É importante observar e entender os diferentes padrões de tráfego para identificar se um website está incorrendo em tráfego humano, de bots básicos ou de bots sofisticados. Esses padrões podem variar entre circadianos, intermitentes e contínuos.

Bots bons vs. bots mal-intencionados

Vamos começar com os aspectos básicos: Um bot, diminutivo de “robot”, é um programa de computador que pode realizar tarefas automatizadas de forma mais rápida e precisa do que um ser humano. Os vários papéis e tipos de bots se enquadram em duas categorias principais: bots bons e bots mal-intencionados (Figura 2). Os bots cinza são uma subcategoria de bots bons, mas vamos mesclá-los com os bots bons por enquanto para simplificar a comparação.

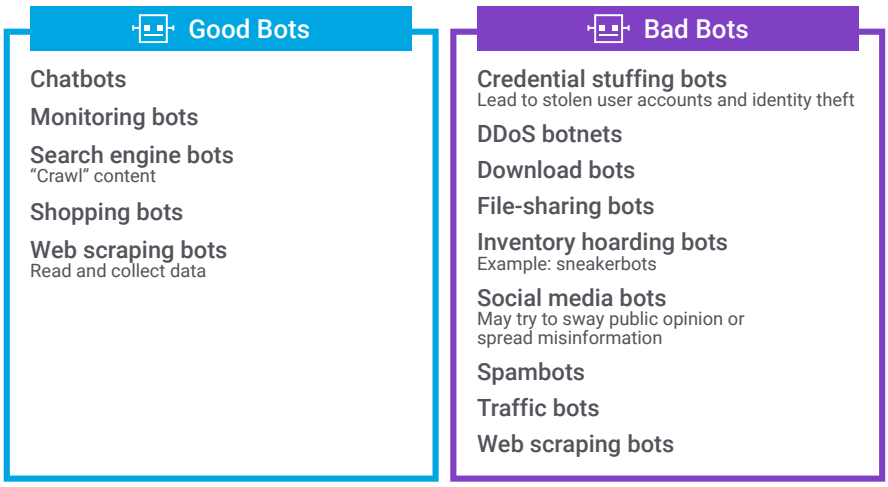
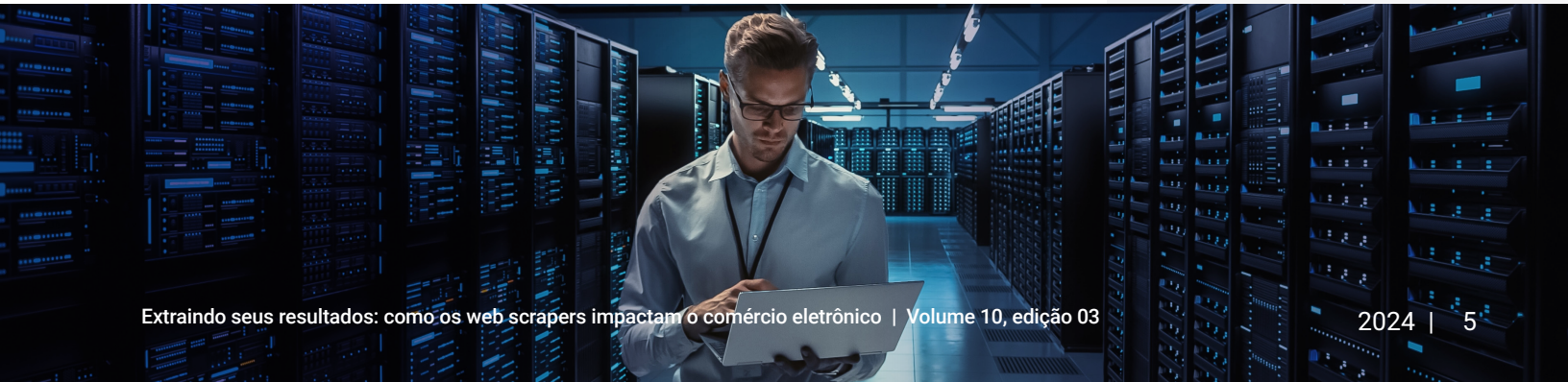


Fig. 2: Uma comparação lado a lado, com exemplos, de bots bons e bots mal-intencionados

Bots bons são bots úteis que ajudam a fornecer ferramentas e serviços, enquanto bots mal-intencionados são frequentemente usados com intenção maliciosa por cibercriminosos e fraudadores. Um exemplo desse tipo de malícia é um bot de tráfego que imita o comportamento humano online para aumentar cliques e tráfego em um website (ou seja, cometer fraude de anúncios).

Bots de web scraping aparecem nas categorias de bots bons e bots mal-intencionados. A distinção tem a ver com a forma como as organizações usam as informações que esses bots coletam. Agora, focaremos mais de perto em vários casos de uso associados aos efeitos bons e ruins dos bots scrapers que são enfrentados por alguns dos maiores varejistas e marcas de comércio eletrônico do mundo.



Conceitos básicos de scraping (“raspagem”)

O web scraping é comumente usado por empresas de comércio eletrônico. Nos setores de viagens e hospitalidade, por exemplo, os agregadores de viagens raspam conteúdo dinâmico de seus hotéis e companhias aéreas parceiros para se manterem atualizados sobre disponibilidade e preços. Esse tipo de raspagem é esperado, e as empresas usam controles comuns de bots para acelerar os scrapers durante os horários do dia em que os usuários reais estão procurando fazer uma reserva. As organizações também usam provedores de serviços de extração de dados para coletar leads e outras informações relacionadas dos concorrentes. Além disso, os bots scrapers podem ser usados para analisar dados e identificar tendências. A raspagem também pode ser benéfica para a análise do website para melhorar as ofertas e serviços online, e para permitir que consumidores potenciais encontrem mais facilmente produtos da empresa, como por meio de um mecanismo de pesquisa. Todas essas ações podem ajudar as empresas a alcançar vantagem competitiva. No entanto, não há como negar que muitas entidades estão usando scrapers por motivos menos recomendáveis.

A raspagem dá uma guinada, e os clientes notam

Infelizmente, muitas vezes ouvimos falar de consumidores que foram vítimas de golpes de phishing. Neste caso, os bots scrapers foram usados para roubar imagens de produtos, descrições e informações sobre preços para criar vitrines falsas ou websites de phishing com o objetivo de roubar credenciais ou informações de cartão de crédito. Esses websites de phishing/falsificação são uma forma de apropriação de marca, na qual a propriedade intelectual das organizações vítimas está sendo usada para estabelecer confiança com clientes em potencial.

Algumas das maiores marcas de comércio eletrônico do mundo foram afetadas por websites falsificados, campanhas de phishing e roubo de dados da Web da empresa como parte das campanhas de apropriação de marca (Figura 3). Infelizmente, quando os websites de phishing são bem-sucedidos, as marcas legítimas sofrem consequências como a perda de confiança e fidelidade do cliente.

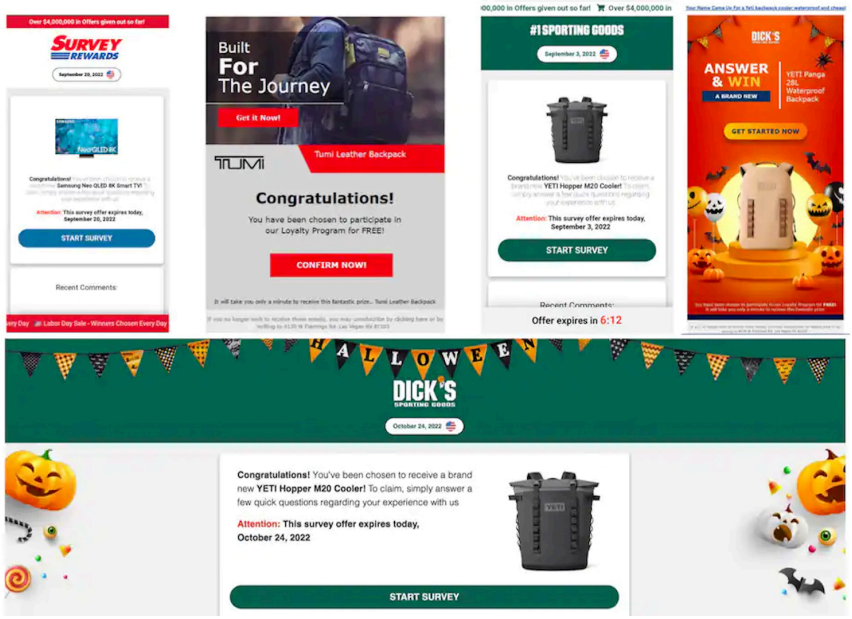


Fig. 3: Um exemplo de algumas das principais empresas de comércio eletrônico que foram vítimas da apropriação da marca

O scalping também pode ser atribuído ao web scraping, pois os scalpers podem raspar um website buscando os produtos disponíveis e comprá-los antes que os clientes legítimos tenham a oportunidade de fazê-lo (Figura 4).

Fig. 4: Casos de uso de scrapers

Os agentes de ameaça que realizam esses tipos de atividades prejudiciais de raspagem estão cientes dos efeitos que seus objetivos mal-intencionados têm sobre as vítimas. Isso inclui os impactos negativos da inteligência/espionagem competitiva, acúmulo/raspagem de inventário, falsificação e criação de websites/mercadorias impostores e raspagem e repostagem de websites de mídia (Tabela 1). E não há leis existentes que proíbam explicitamente o uso de bots scrapers.

Impacto	Descrição
Inteligência competitiva/ espionagem	Os concorrentes usam as informações do website de uma organização para reduzir os preços, fazer alterações em suas ofertas e ter uma noção de novas oportunidades e ameaças.
Acúmulo/raspagem de inventário	Os scalpers fazem buscas constantemente em websites direcionados para encontrar produtos que se tornam disponíveis e, em seguida, adicioná-los a carrinhos, tornando esses produtos indisponíveis para clientes reais.
Falsificação e websites/ mercadorias impostores	Os falsificadores usam conteúdo raspado para criar websites e catálogos de produtos falsos para induzir os usuários a pensar que estão comprando produtos legítimos em vez de falsificações.
Roubo e repostagem de websites de mídia	Os invasores podem raspar artigos de notícias, blogs e outros conteúdos e colocá-los em seus próprios websites, fazendo com que a organização original perca os visitantes e potencial receita de anúncios. As taxas de publicidade são frequentemente baseadas no número de visitantes/público no website, de forma que menos visitantes significam que o website da mídia perde a receita que obteve de taxas de anúncio mais elevadas.

Tabela 1: Impactos negativos intencionais causados por web scrapers

Os efeitos colaterais gerais do web scraping

Independentemente da intenção do web scraping, as organizações precisam lidar com as despesas de seus efeitos colaterais. Algumas empresas pagam por serviços de raspagem benéficos, mas as empresas que estão sendo raspadas estão incorrendo em custos próprios. Isso inclui despesas com soluções antibot e os impactos econômicos negativos da degradação do desempenho do website e da poluição das métricas principais (Tabela 2).

Impacto	Descrição
Aumento dos custos de servidor, CDN (Rede de Entrega de Conteúdo) e nuvem para atender ao tráfego de bots	Isso afeta a receita e causa perda de reputação decorrente do uso de conteúdo por parte dos concorrentes, invasores e falsificadores.
Degradação do desempenho do website	Como os bots scrapers são executados continuamente até serem interrompidos, esses bots aumentam os custos de servidor e entrega conforme as organizações atendem ao tráfego indesejado de bots e sofrem de experiências de usuário comprometidas, como desempenho mais lento de websites e apps.
Poluição das métricas principais	A atividade de bots não detectados distorce gravemente as principais métricas, como a conversão de websites, das quais as equipes de negócios dependem para tomar decisões de investimento, como estratégias de posicionamento de produto e campanhas de marketing.

Tabela 2: Impactos negativos não intencionais causados por web scrapers

Raspagem para contratação: serviços de web scraping de terceiros

Como mencionamos, os bots web scrapers podem ser usados para o bem e para o mal. Ao contrário dos bots usados para ataques de preenchimento de credenciais, que são bots mal-intencionados conhecidos e, portanto, justificadamente bloqueados, existem empresas que oferecem bots de web scraping legítimos. Muitas organizações usam esses serviços de web scraping de terceiros para extrair e fornecer dados para sua própria organização, o que pode ser benéfico, especialmente no mundo do marketing competitivo.

Existem dezenas dessas empresas que fornecem diferentes tipos de serviços de extração de dados/web scraping; há até mesmo conferências que as promovem. Por exemplo, a Bright Data hospeda uma conferência chamada ScrapeCon que reúne especialistas em evadir detecções de bots para que as empresas possam aprender a raspar dados. A Tabela 3 inclui exemplos dos níveis de serviços que podem ser fornecidos por empresas de web scraping de terceiros.

Nível de serviço 1	Os serviços de proxy podem fazer parte da raspagem e oferecem uma infraestrutura que pode incluir endereços IP móveis e residenciais dos data centers.
Nível de serviço 2	Esse segundo nível também pode incluir a extração automatizada de dados que limpa e estrutura os dados para facilitar o uso pelos membros da equipe de ciência de dados do cliente, que extraem a valiosa inteligência para orientar as decisões de negócios.
Nível de serviço 3	O nível mais alto pode adicionar a extração da inteligência de negócios real em si, o que pode melhorar ainda mais o processo de tomada de decisão para as empresas. Eles são chamados de “botnets de IA”.

Tabela 3: Vários níveis de serviços fornecidos por empresas de web scraping de terceiros

Os clientes podem escolher qualquer um desses níveis de serviço, do mais básico ao mais avançado, bem como a frequência da coleta de dados, e podem especificar seus destinos. Muitas vezes, o nível de serviço fornecido, ou botnet escolhido, depende do nível de proteção que eles precisam superar. Um botnet mais básico pode coletar dados por meio de um script avançado com alguns milhares de servidores proxy localizados em data centers que equilibram a carga de tráfego. Se a proteção for rudimentar o suficiente, o botnet poderá usar essa técnica para passar pelas defesas de gerenciamento de bots e pelo firewall de aplicativos da Web da infraestrutura de segurança.

Se, no entanto, a proteção for mais avançada, poderá ser necessária uma abordagem mais sofisticada de raspagem, como um ataque de navegador headless. Isso se aplica tanto para a raspagem conduzida por um agente com boa intenção quanto para um com má intenção. E não é barato, pois as empresas incorrerão em custos que geralmente são muito mais altos para a infraestrutura mais sofisticada do que para o nível básico de serviço. Uma defesa avançada pode incluir tecnologias de desafio (como CAPTCHA ou prova de trabalho), várias camadas de detecção projetadas para avaliação de impressão digital no lado do cliente e uma análise das características do HTTP (Protocolo de transferência de hipertexto) e TLS (Segurança de camada de transporte).

O processo de raspagem para botnets de IA

Embora os web scrapers básicos possam ser mais consistentes em suas técnicas de raspagem, os botnets de IA têm a capacidade de descobrir e de raspar conteúdo e dados não estruturados que estejam em um formato ou local menos consistente. Além disso, os botnets de IA podem usar a inteligência de negócios real para melhorar o processo de tomada de decisões. Os botnets de IA sofisticados, mencionados na Tabela 3, nível de serviço 3, têm um processo de raspagem de dados em três etapas. Eles operam coletando, extraindo e processando dados (Figura 5).

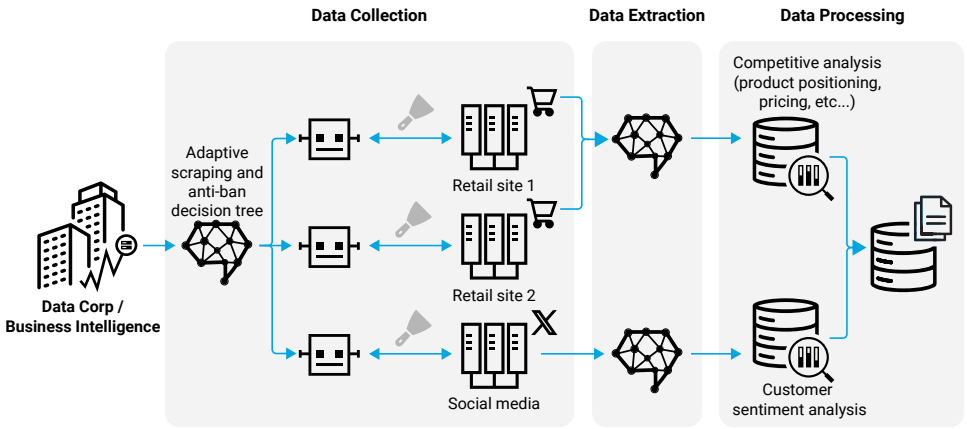


Fig. 5: Uma representação de um botnet de IA e seu processo de três etapas

Vamos examinar essas três etapas em mais profundidade para entender melhor o que elas comportam.

Coleta de dados

O web scraping envolve a organização de dados que foram extraídos de um website, ou de websites, para que as organizações possam produzir novos conjuntos de dados que possam ser aplicados e analisados da maneira mais conveniente. E isso começa com a coleta dos dados.

É necessário que a coleta de dados seja composta por raspagem adaptável, combinada com tecnologias “antiban” ou “detecção antibots”, para funcionar rapidamente e sem problemas. Essas tecnologias são configuradas como árvores de decisão para detectar vários aspectos de qualquer proteção que possa estar instalada. Resiliência é o nome do jogo aqui. A proteção contra bots pode incluir impressão digital JavaScript, impressão digital HTTP e TLS (avaliando os cabeçalhos HTTP e handshake de TLS) e detecção de reputação de IP (Protocolo de Internet) (Figura 6). Alguns desses fluxos de trabalho podem incluir ML (machine learning), especialmente ao coletar estatísticas sobre a taxa de sucesso; ajuste à estratégia de cookies, cabeçalho HTTP e parâmetros TLS; e avaliação do código de impressão digital JavaScript. É também aqui que um navegador headless pode entrar em jogo.

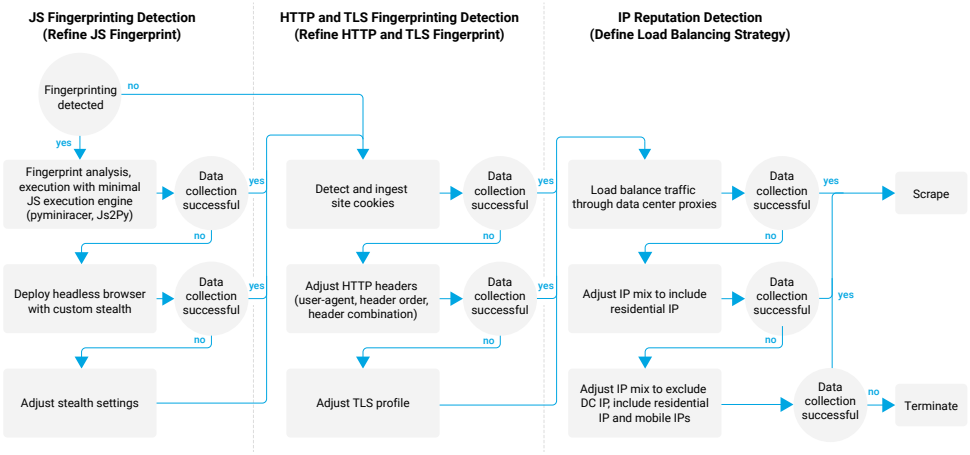


Fig. 6: Ao tentar coletar dados, essa árvore de decisão de detecção antibot tenta evitar a impressão digital de JavaScript, a impressão digital de HTTP e TLS e a detecção de reputação de IP

O navegador sem cabeça (headless)

Um navegador headless é um navegador da Web que não tem uma interface do usuário gráfica (GUI). Isso significa que os seres humanos não podem interagir diretamente com a página da Web na qual o navegador headless aparece, e o navegador é executado por meio de uma CLI (interface de linha de comando) ou por uma comunicação de rede. No caso do Selenium, um popular navegador headless de código aberto, ele é automatizado e amplamente usado para web scraping. Isso pode ser muito útil para aqueles que buscam dados que estão tentando raspar conteúdo dinâmico.

Os navegadores headless também podem permitir que as capturas de tela e o código do website sejam copiados com eficiência, e que os dados escolhidos sejam extraídos sem renderizar a página inteira. No entanto, os ataques a navegadores headless são caros de conduzir e, às vezes, podem ser detectados pelas impressões digitais que deixam para trás. No entanto, as despesas de outras infraestruturas sofisticadas são semelhantes às de navegadores headless, ou seja, geralmente altas.

Extração de dados e processamento de dados

As informações extraídas geralmente consistem em conteúdo HTML e JSON. De todos os dados extraídos, apenas uma parte pode ser útil para a análise. Por exemplo, a análise competitiva geralmente inclui preços, descontos, inventário e números de SKU de produtos, categorias e descrições. Partes essenciais de informações podem ser extraídas automaticamente por modelos de ML que podem ser treinados com várias estruturas e formatos de dados para reconhecê-las. Isso ajuda a evitar todo o trabalho de processamento extra que deve ser feito para extrair manualmente os dados e ajuda a evitar a necessidade de estudar a estrutura de código de conteúdo HTML e JSON. Além disso, a estrutura do código de conteúdo pode mudar à medida que o design do website evolui. A lógica de ML adicional também é necessária para o processamento se vários websites estiverem envolvidos como parte do escopo da análise.

Estudo de caso: benefícios das soluções de detecção do web scraping

Os pesquisadores da Akamai observaram um subconjunto de clientes de comércio eletrônico que estavam protegidos por uma solução de web scraping que detectava atividades de raspagem e analisaram o detalhamento da atividade de tráfego por uma semana. Isso totalizou uma amostra de aproximadamente 6,9 bilhões de solicitações. A análise levou em conta apenas as solicitações de HTML e AJAX. O conteúdo estático (imagens, JavaScript, folhas de estilo) não foi incluído na análise, pois a maioria dos bots não solicita conteúdo estático; essa omissão também ajudou a evitar o enchimento desnecessário dos dados.

A atividade geral foi classificada pelo Akamai Content Protector e consistiu em 49,3% de tráfego humano de baixo risco, 42,1% de tráfego de bots (27,5% de bots mal-intencionados de alto risco e 14,6% de bots bons) e 8,7% de tráfego não classificado de risco médio (Figura 7).

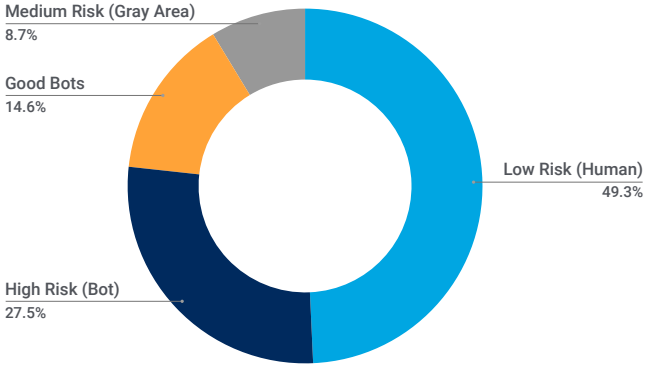


Fig. 7: Detalhamento da classificação de atividade de tráfego

A Figura 8 mostra que, dos 42,1% do tráfego de bots, 65,3% se originaram de scrapers considerados bots mal-intencionados, e os 34,7% restantes eram de scrapers classificados como bots bons (por exemplo, mecanismos de pesquisa na Web, SEO, mídia social e publicidade online).

Fig. 8: Tráfego de bots bons vs. tráfego de bots mal-intencionados

Os níveis de sofisticação também foram medidos para os bots mal-intencionados de alto risco que contribuíram para 65,3% do tráfego geral de bots. Do total, 37% desse tráfego vieram de botnets básicos com script que são fáceis de detectar por meio de métodos simples sem estado, 47,6% vieram de botnets com script mais avançados que exigem métodos mais avançados de detecção com informações de estado usando ML, e 15,5% vieram de navegadores headless que exigem métodos avançados de impressão digital em JavaScript e detecção com estado (Figura 9).

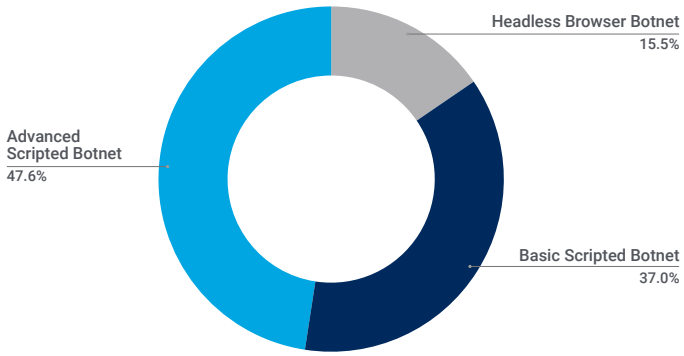


Fig. 9: Distribuição de tráfego de bots mal-intencionados com base em sua sofisticação (os totais não somam 100% devido ao arredondamento)

Assim, a partir desses dados, podemos ver que os scrapers de bots mal-intencionados são significativamente mais numerosos do que os scrapers de bots bons, e que quase metade do tráfego geral consistia em bots, com os botnets com script avançados produzindo o maior tráfego de bots mal-intencionados (47,6%).

A atividade do website será muito mais rápida e eficiente, e as métricas do website serão mais limpas para ler, uma vez que as defesas contra esses bots estejam em vigor e os scrapers sejam removidos. E esses resultados proporcionarão em melhores experiências do usuário/cliente. Conforme mostrado na Figura 10, o número de solicitações de bots de alto risco diminuiu substancialmente quando a mitigação foi ativada.

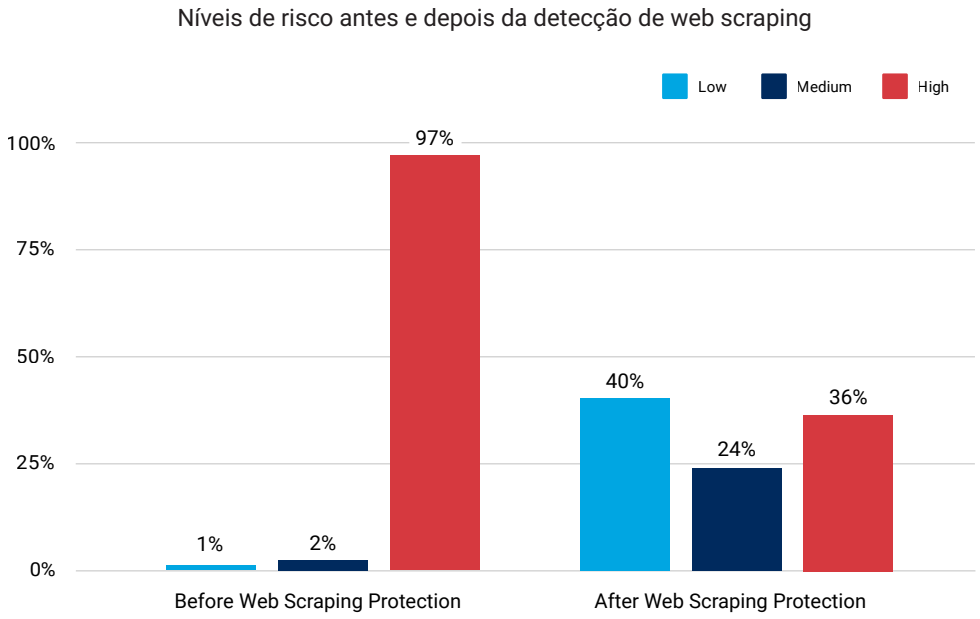


Fig. 10: Níveis de risco antes e depois da mitigação com o Content Protector

Proteção e mitigação

Esta seção fornece alguns indicadores cruciais na detecção de web scrapers e informações sobre ferramentas que podem fornecer medidas defensivas contra eles.

Detecção de scrapers básicos

Embora os scrapers sofisticados possam ser difíceis de detectar, as soluções de gerenciamento de bots podem defender contra a coleta de dados por todos os tipos de scrapers intrusivos e podem, especialmente, procurar as seguintes caraterísticas para detectar bots web scrapers mais simples:

- Solicitações que anunciam versões de SO e navegadores mais antigos
- Anomalias na assinatura do cabeçalho HTTP
- O uso de versões antigas do HTTP (por exemplo, v1.1) em vez do HTTP v2 mais comum ou do HTTP v3 emergente
- Solicitações provenientes de milhares de serviços de nuvem/data centers

Detecção de scrapers mais avançados

Nenhuma das caraterísticas na lista acima será observável para os scrapers mais avançados. Confira aqui algumas caraterísticas dos scrapers mais sofisticados:

- Solicitações provenientes do navegador e da versão do SO mais recentes
- O conjunto de cabeçalhos HTTP parece idêntico ao navegador legítimo
- O uso de HTTP v2
- Solicitações provenientes de centenas de milhares de endereços IP móveis e residenciais

Identificação de padrões de tráfego

Há alguns indicadores-chave que podem identificar se o tipo de tráfego que um website está incorrendo é humano (Figura 11), bot básico (Figura 12) ou bot sofisticado (Figura 13).

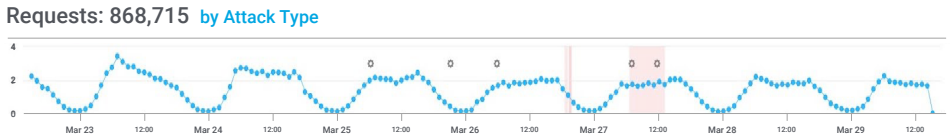


Fig. 11: O tráfego legítimo do usuário geralmente mostra um ciclo circadiano de atividade

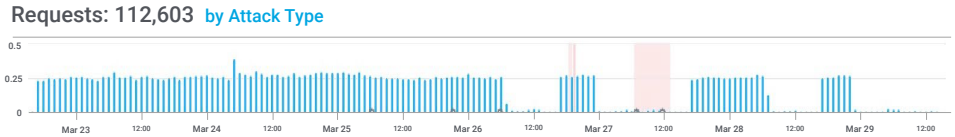


Fig. 12: O tráfego típico de bots exibe atividades regulares com pausas ocasionais

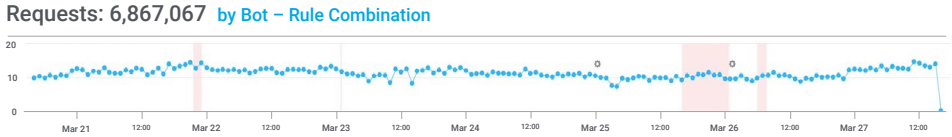


Fig. 13: Bots mais sofisticados mostram o tráfego continuamente, dia e noite

Muitas vezes também vemos botnets que estão no meio termo, com uma estratégia de balanceamento de carga fraca, mas uma estratégia sofisticada de impressão digital (ou vice-versa). No entanto, botnets mais avançados podem ser tão sofisticados que podem passar como tendo uma impressão digital perfeita ou até mesmo reproduzir um padrão legítimo de tráfego de usuários.

Além de estar atento a esses bots scrapers, ferramentas que protegem contra o web scraping, como um protetor de conteúdo, podem possibilitar benefícios especiais e uma navegação mais tranquila entre as águas infestadas por scrapers. Os benefícios podem incluir:

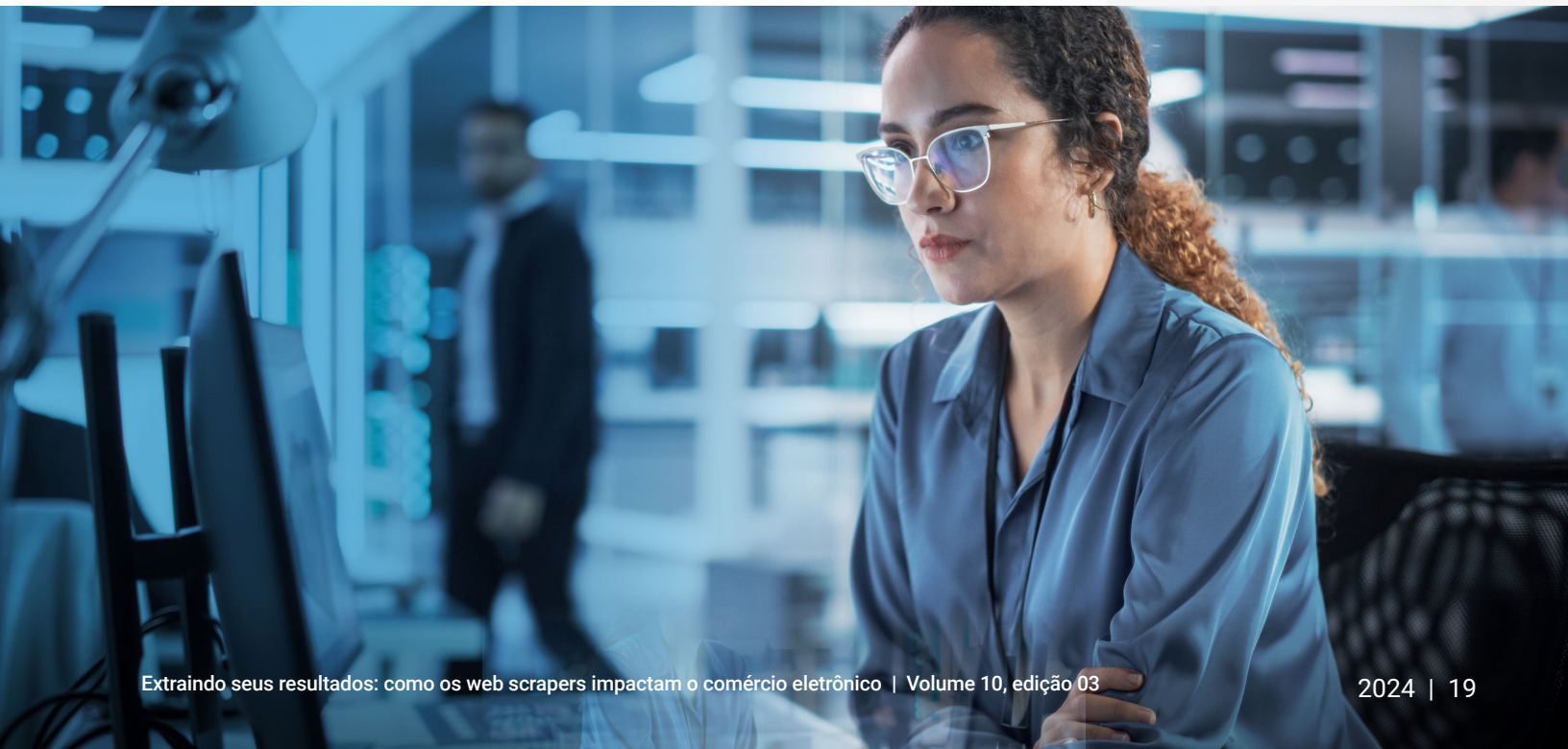
- Taxas de conversão mais altas e custos de TI reduzidos
- Métricas mais precisas, que podem levar a melhores decisões de investimento e aumentar a receita
- Pressão de preços reduzida, que pode se traduzir em vendas salvas da redução de preços do concorrente
- Clientes satisfeitos que podem acessar os produtos desejados, além de aumentar a receita de oportunidades de upsell quando os clientes adicionam outros produtos ao carrinho depois de terem garantido o item principal
- A reputação preservada da marca, pois os clientes são protegidos contra falsificações de baixa qualidade que eles acham que são produtos legítimos do vendedor original
- Retenção da receita do produto e manutenção da fidelidade do cliente
- Aumento/proteção da receita de anúncios
- Retenção de público e visitantes do website

Considerações de conformidade

O PCI DSS v4.0 (Padrão de segurança de dados do setor de cartões de pagamento) já está em vigor, e muitas das mudanças foram impulsionadas por tendências de ameaças que ainda estão tendo um impacto sobre as empresas. A visibilidade é fundamental para lidar com esses ataques. Quer estejam em seu ambiente JavaScript histórico ou em APIs usadas para facilitar a transformação, é essencial detectar e corrigir rapidamente esses ataques.

Também vemos tendências emergentes de conformidade no novo NIST Cybersecurity Framework versão 2.0, que adicionou uma função de governança. O NIST tende a ser uma base para uma série de normas governamentais e afeta muitas estruturas de cibersegurança comerciais. Portanto, agora é um ótimo momento para analisar a nova orientação e usá-la para atualizar suas políticas ou mapear sua documentação atual para ver onde você tem lacunas.

Para empresas de capital aberto e aquelas que usam princípios contábeis geralmente aceitos (GAAP), outra área de conformidade é a materialidade da cibersegurança. A necessidade de definir riscos e ameaças relevantes requer colaboração em toda a equipe de liderança. Depois de identificar ameaças relevantes (como ransomware), você precisa mapear mitigações (como microssegmentação). Certifique-se de que seus planos de gerenciamento de crises atendam aos cronogramas de divulgação e tenha um manual que contemple o pior cenário, no qual você precisaria registrar um Formulário de Incidentes Cibernéticos 8-K da Security and Exchange Commission (Comissão de Valores Mobiliários).



Conclusão

Esperamos que este relatório lhe dê uma visão de uma área que possa estar sofrendo impactos econômicos negativos em sua organização. Os bots estão afetando seus websites em volumes cada vez maiores, e é importante otimizar os bots benéficos, mitigar bots mal-intencionados e garantir baixo atrito na experiência geral do cliente. Esse é um problema de segurança com impactos nos negócios. Assim como ocorre com todos os problemas de segurança, a primeira etapa é ganhar visibilidade, a segunda é analisar o impacto e a última etapa é determinar o ROI de risco e receita para que você possa implementar os controles de segurança apropriados.

Não é possível proteger o que não é possível ver, por isso agora é o momento de determinar onde existem lacunas na visibilidade. Para fazer isso, você deve determinar o nível de atividade de web scraping em seus websites e sua intenção. Tanto bots bons quanto bots mal-intencionados compõem o cenário de bots, e os bots scrapers estão em ambas as categorias, dependendo do uso. Embora a linha entre bots scrapers benéficos e prejudiciais possa ser pouco nítida, a evolução da sofisticação dos bots (por exemplo, web scrapers conduzindo ataques de navegador headless) continua. Tudo isso vem com o enorme impacto que os bots web scrapers têm entre as entidades de comércio eletrônico nos custos de TI e na experiência do cliente. É fundamental garantir que você tenha as ferramentas em funcionamento para analisar a atividade de bot e os impactos em seu website.

O que você não quer são os invasores que executam seu modelo de negócios criminoso em seus websites e cometem diversas atividades mal-intencionadas, como eliminar pontos de fidelidade, fazer pedidos fraudulentos ou até mesmo conduzir fraudes de devolução. Você também não quer que os bots de ingresso esgotem eventos limitados ou que os bots de compras comprem produtos em alta. Os bots podem ser usados para facilitar o abuso de abertura de novas contas, aproveitando ofertas especiais, o que afeta a análise e os custos da campanha. Grandes botnets de DDoS (Negação de serviço distribuído) podem sobrecarregar aplicativos voltados para a Web e causar uma experiência de usuário insatisfatória ou a incapacidade de fazer pedidos ou reservas, resultando em perda de receita e atrito com os clientes. Os bots podem até mesmo imitar o comportamento humano online para aumentar os cliques e o tráfego em um website, distorcendo tanto a análise de marketing quanto a de desempenho de experiências digitais cuidadosamente criadas. Você definitivamente não quer nada disso.

Como observamos anteriormente, mais da metade do tráfego global de comércio na Web é composta de bots, e os níveis de tráfego de bots continuam a aumentar. A Akamai baseou os insights e conselhos neste relatório em nossa plataforma de segurança, que inclui proteção de conteúdo com defesa contra web scraping. Estabelecemos parcerias com muitos líderes de comércio eletrônico, por isso queríamos compartilhar proteções e mitigações que as empresas possam usar para proteger melhor seus clientes. Prevemos um aumento no uso, nas opções de nível de serviço e nos tipos de bots web scrapers disponíveis. Portanto, é necessário avaliar continuamente a postura da sua empresa em relação aos riscos e determinar se os controles de segurança atuais estão atendendo ao apetite por riscos da sua liderança.

Fique por dentro de nossas pesquisas mais recentes conferindo nosso hub de pesquisa de segurança.

Metodologias

Dados do Content Protector

Esta amostra de dados descreve as classificações de nível de risco que nossa ferramenta Content Protector atribui ao tráfego que ela monitora. Essas classificações são usadas para detectar atividades de bot scraping e para determinar se estamos lidando com um bot bom ou mal-intencionado. Como a maioria dos bots não solicita conteúdo estático, essa análise levou em conta apenas as solicitações de HTML e AJAX para evitar o enchimento desnecessário dos dados.

Esta amostra de dados abrangeu o período de uma semana de 12 a 19 de abril de 2024. Nosso tamanho total da amostra consistiu em mais de 6,5 bilhões de solicitações.

Ataques de bots

Esses dados descrevem alertas da camada de aplicativos sobre o tráfego visto por meio de nosso WAF (firewall de aplicativos da Web) e nossa ferramenta de gerenciamento de bots. Os alertas de bot são acionados quando detectamos uma carga útil de bot em uma solicitação enviada a websites, aplicativos ou APIs protegidos. Esses alertas de bots podem ser acionados por bots maliciosos e benignos. Os alertas não indicam o êxito de um ataque. Embora esses produtos permitam um alto nível de personalização, coletamos os dados apresentados aqui de uma forma que não considera as configurações personalizadas das propriedades protegidas. Os dados foram extraídos de uma ferramenta interna para análise de eventos de segurança detectados na Akamai Connected Cloud, uma rede de aproximadamente 340 mil servidores em mais de 4 mil locais em quase 1.300 redes em mais de 130 países. Nossas equipes de segurança usam esses dados, medidos em petabytes por mês, para pesquisar ataques, sinalizar comportamentos mal-intencionados e apresentar inteligência adicional às soluções da Akamai.

Esses dados cobriram um período de 15 meses de 1º de janeiro de 2023 a 31 de março de 2024.

Créditos

Editor-chefe

Lance Rhodes

Editorial e redação

David Senecal
Maria Vlasak

Análise e contribuição do assunto

Mitch Mayne
Susan McReynolds
Christine Ross
Badette Tribbey
Steve Winterfeld

Análise de dados

Chelsea Tuttle

Materiais promocionais

Annie Brunholzl

Marketing e publicação

Georgina Morales
Emily Spinks

Mais informações sobre o State of the Internet/Security

Leia as edições anteriores e fique por dentro das próximas versões dos aclamados relatórios State of the Internet/Security da Akamai. **akamai.com/soti**

Mais informações sobre a pesquisa de ameaças da Akamai

Mantenha-se atualizado com as mais recentes análises de inteligência de ameaças, relatórios de segurança e pesquisas sobre cibersegurança. **akamai.com/security-research/**

Acesse os dados deste relatório

Visualize versões em alta qualidade das tabelas e dos gráficos mencionados neste relatório. Essas imagens podem ser usadas e consultadas livremente, desde que a Akamai seja devidamente creditada como a fonte e que o logotipo da Akamai seja mantido. **akamai.com/sotidata**

Saiba mais sobre as soluções da Akamai

Para saber mais informações sobre as soluções da Akamai para detecção e proteção contra web scrapers, visite nossa **página do Content Protector**.

A Akamai protege sua experiência do cliente, sua força de trabalho, seus sistemas e seus dados ajudando a incorporar a segurança em tudo o que você criar, em qualquer lugar que você criar e entregar. A visibilidade da nossa plataforma em relação às ameaças globais nos ajuda a adaptar e desenvolver sua postura de segurança para permitir Zero Trust, interromper ransomware, proteger aplicativos e APIs ou combater ataques DDoS, dando a você a confiança para inovar, expandir e transformar continuamente o que é possível. Saiba mais sobre as soluções de computação em nuvem, segurança e entrega de conteúdo em akamai.com e akamai.com/blog, ou siga a Akamai Technologies no X, antigo Twitter, e no LinkedIn.

Publicado em 06/24.